PDB4DNA: Implementation of DNA geometry from the Protein Data Bank (PDB) description for Geant4-DNA Monte-Carlo simulations

E. Delage, Quang Trung Pham Pham, M Karamitros, H Payno, V Stepan, S Incerti, L. Maigne, Y Perrot

PDB4DNA: implementation of DNA geometry from the Protein Data Bank (PDB) description for Geant4-DNA Monte-Carlo simulations

E. Delage[a], Q.T. Pham[a], M. Karamitros[b], H. Payno[a], V. Stepan[b], S. Incerti[b], L. Maigne[a], Y. Perrot[a]

[a] Clermont Université, CNRS/IN2P3, Laboratoire de Physique Corpusculaire de Clermont-Ferrand, LPC, 24 avenue des Landais, BP 80026, 63171 Aubière, France

[b] Université Bordeaux, CNRS/IN2P3, Centre d'Etudes Nucléaires de Bordeaux-Gradignan, CENBG, Chemin du Solarium, BP 120, 33175 Gradignan, France

Corresponding author: Yann Perrot

tel: +33 4 73 40 78 49

fax: +33 4 73 26 45 98

perrot@clermont.in2p3.fr

Abstract

This paper describes PDB4DNA, a new Geant4 user application, based on an independent, cross-platform, free and open source C++ library, so-called PDBlib, which enables use of atomic level description of DNA molecule in Geant4 Monte Carlo particle transport simulations. For the evaluation of direct damage induced on the DNA molecule by ionizing particles, the application makes use of an algorithm able to determine the closest atom in the DNA molecule to energy depositions. Both the PDB4DNA application and the PDBlib library are available as free and open source under the Geant4 license.

Program summary

Title: PDB4DNA

URL: http://pdb4dna.in2p3.fr

Licensing provisions: Geant4 license, http://geant4.web.cern.ch/geant4/license/LICENSE.html

Programming language: C++

Operating system: cross-platform

1. Introduction

The simulation of radiation-induced effects in living tissues requires the modeling of different action mechanisms [1]: the physical stage describing the elementary physical interactions causing

ionization and excitation of water molecules; the physico-chemical and chemical stages describing the production, the diffusion and the reaction of chemical species and the biological stage predicting radiation effects in living tissues. The modeling of physical, physico-chemical and chemical actions has to be coupled with a geometrical model describing correctly the targets at the origin of the biological endpoints. It has been demonstrated that damage to nuclear DNA are responsible for biological effects such as chromosome aberrations, mutations and cell inactivation [2,3]. Consequently, the correct modeling of cell nucleus content is required in order to estimate the detailed clustering of energy deposition at this scale. Modeling the three dimensional structure of the DNA molecules is a way to describe the cell nucleus. Indeed, the precise implementation of DNA geometry is of key importance in order to take into account the spatial structure of energy depositions along the particle track generated by Monte Carlo Track Structure (MCTS) codes. However, the modeling of DNA geometries is not trivial if we keep in mind that DNA molecule for a normal human cell represents about 3.1×10^9 base pairs [4] geometrically arranged in many structural levels that evolve with time and may vary between cells of a same line. Among the variety of MCTS codes [5], the strategy consists in replicating and arranging an elementary piece of DNA to form structural levels. Usually, each elementary piece of DNA is represented either by a simple geometrical shape [6,7,8,9,10,11] or by a piece of DNA double helix in atomic resolution [1,12,13,14,15].

Historically, geometrical DNA targets were treated as small cylinders in order to compute energy depositions comparable with microdosimetry or nanodosimetry experiments [7,8,16]. To allow the computation of DNA strand breaks, small cylinders were divided to reach the definition of DNA bases and sugar phosphate backbones [6]. Bernal and Liendo proposed a more structured geometrical model for DNA molecules [9]. Assuming that energy depositions in sugar-phosphate groups cause DNA damage, they were represented as prisms and arranged using parametric equations to enable computations up to 1.2×10^8 base pairs. This method guaranteed the packing ratio of the B-DNA structure taken from literature [17]. This geometrical model, first implemented to explore the capacities of the PENELOPE code to calculate DNA strand breaks [9], was adapted to Geant4 to study the production of DNA strand breaks [10,18]. More recently, Dos Santos *et al.* implemented a geometrical model using Geant4 for which spheres representing bases and sugar phosphate backbone

were replicated using parameterized geometrical transformations [11]. Compared to the work of Bernal and Liendo, the model of Dos Santos *et al.* was able to describe the whole DNA content of specific human cell nuclei (endothelium and fibroblast in the G0/G1 phase) and was proposed as the so-called "dnageometry" Geant4 advanced example.

More sophisticated geometrical models where DNA is represented by detailed atomic structure can be found in the literature but are not available publicly. Michalik and Begusova defined a geometrical non-homogeneous target model of a nucleosome to be used with the TRION MCTS code [12]. To perform this geometrical model, a linear piece of DNA was wrapped around the histone core represented by a cylinder. Nikjoo and Girard proposed geometrical models [13], implemented in PHITS and KURBUC MCTS codes, for which 30 nm solenoid describing chromatin fibers are derived from a canonical B-DNA structure and are then used to fill chromosome domains to construct the complete cell genome. In a similar way, the PARTRAC (PARticle TRACks) code is able to describe the whole genome of specific human cells (fibroblast or lymphocyte in the G0/G1 phase) [1,14]. In this case, the cell genome is defined as a DNA double helix target detailed up to the atomic scale. Five straight fiber elements and four bent elements build up a random-walk fiber model on a grid of 50 nm x 50 nm. Thus, the main advantage is that chromatin fiber is defined by a flexible arrangement of nucleosomes allowing an unbroken DNA segment.

It has to be noticed that these sophisticated geometrical models are implemented in specific purpose MCTS codes that are not publicly available. Therefore, in order to provide to a large community of users advanced geometry of DNA molecules, we used in this study the general purpose Monte Carlo toolkit Geant4 [19]. This toolkit is being extended to handle microdosimetry and radiobiology applications in the framework of the Geant4-DNA project [20] by the inclusion of detailed physics models [21], radiochemistry simulation [22,23] and DNA geometrical models [11]. Geant4 advanced features to describe, place and replicate basic geometries can be used to describe directly DNA structure.

Bernal *et al.* [24] proposed a formalism, based on an atomic-resolution geometrical model of the B-DNA configuration, that was tested with Geant4 but unfortunately no user example was released.

The authors suggested that "for DNA-radiation interaction simulations, a formalism and an associated code to determine the closest atom to an arbitrary point in space is needed".

The need to model, store and exchange polyatomic structures has driven the emergence of file formats enabling the publication, exchange and reuse of data. The Protein Data Bank (PDB) file format [25] represents the molecular geometry by defining the relative position of the constituting atoms. In the PDB file format, no information relative to electronic occupancy are included. The structure can be either obtained experimentally by crystallography techniques or computed with geometry optimization codes which are often able to directly generate PDB files.

The aim of this paper is to describe a Geant4 user application called PDB4DNA that simulates energy deposition in a target volume generated from a PDB file representing the geometry of DNA molecule and estimates energy depositions in such a geometry. PDB4DNA is the first initiative, written in C++, free and open source, integrating the description of molecular geometry from PDB in Geant4 simulations in order to give an estimation of the strand breaks occurring in DNA geometry. Geometry extracted from PDB allows the implementation of the most realistic geometrical description of macromolecules in Geant4 ever performed for DNA damage.

In a first part, for clarity purpose, we present the UML diagram of the application and a description of a PDB file structure. Then, we provide details about functionalities on which the application is based. Then, we present the implementation of the molecular geometry under Geant4 and the resulting visualization by reproducing the exact position, as stored in a PDB file, of each atom represented by a sphere of Van der Waals radius. Finally, we propose an algorithm for assigning the energy deposition locations to DNA geometry at the atomic scale. This feature is used by the PDB4DNA user application to give an estimation of direct DNA damage such as single strand and double strand breaks.

2. PDB4DNA user application

PDB4DNA is a Geant4 application that reads any PDB file. It parses the data and uses them to create corresponding Geant4 description of the geometry, assigns energy depositions to atoms constituting DNA molecule in order to estimate direct damage such as strand breaks. Even though PDB files describe geometry for large molecules (protein, nucleic acids), we focus on DNA geometrical

information for the simulation. In this work, simulations based on other types of molecules are not handled. Users can also visualize any molecule and particle tracks thanks to Geant4 visualization drivers.

Fig. 1 presents the UML diagram of the PDB4DNA Geant4 example. The group of classes in white corresponds to the standard virtual classes that Geant4 users may implement. The core of the PDB4DNA application is represented by the group of classes in green. For easy distribution, all the features independent of Geant4 have been grouped into a C++ library called *PDBlib*, represented by the group of classes in blue. In the following sub-parts we discuss the different functionalities of this library.

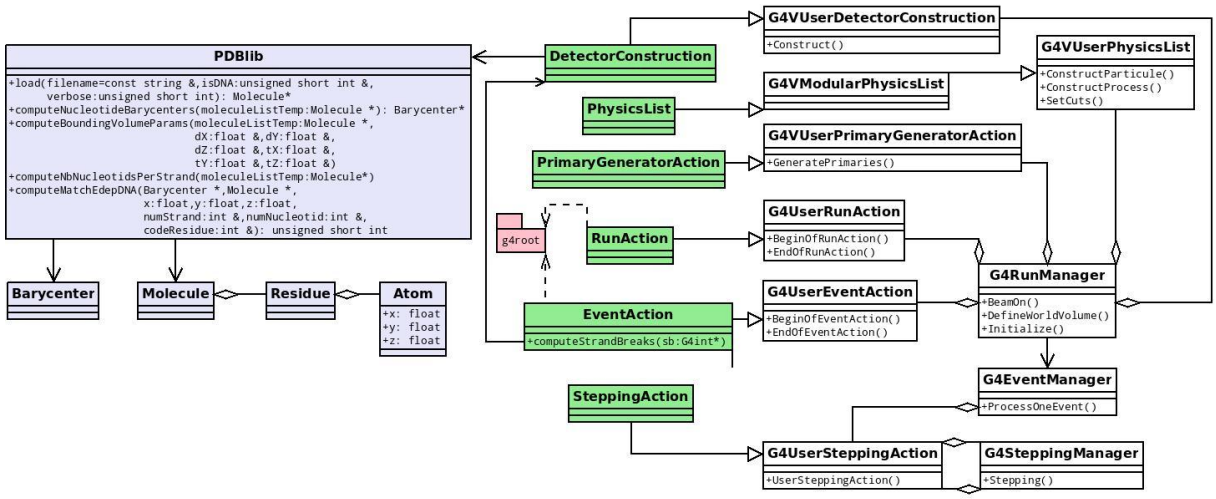


Fig. 1 : UML diagram of the PDB4DNA Geant4 user application: Geant4 virtual classes (white), Geant4 implemented classes (green), PDB library (purple) and interface to ROOT analysis software (red).

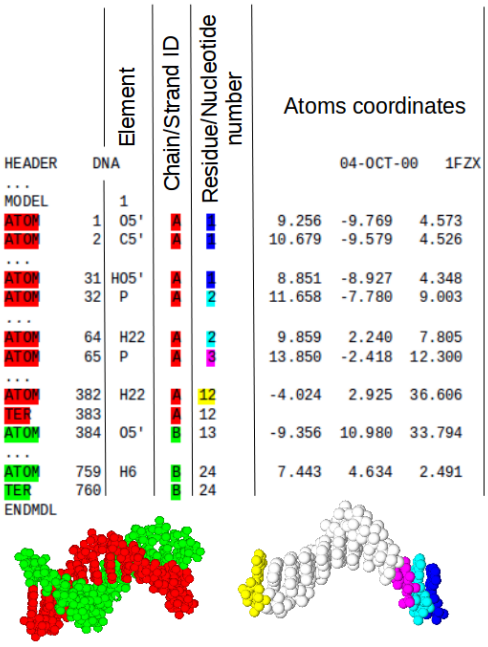


Fig. 2 : PDB file format description. This format provides atom coordinates for each nucleotide number (dark blue, light blue, purple, yellow) constituting each DNA strand (red and green).

2.1. PDB reader

The *PDB reader* functionality (called by the function *PDBlib::load*, Fig. 1) reads the molecular structure from a PDB file into Geant4. Note that the *PDB reader* is able to read and convert any other macro molecules into Geant4 geometry, such as proteins. In order to propose a lightweight toolkit fully integrated to a Geant4 user application, our *PDB reader* has been fully implemented while existing C++ libraries such as "A simple C++ PDB reader" [26], "Easy Structural Biology Template Library" [27], "OpenBabel"[28], were considered feature creep for a Geant4 user application. As a consequence, our Geant4 user application is ready to use without any third-party library. As shown in Fig. 2, a PDB file is an ASCII file in which each line is describing the structural information of the molecule. Each line is 80 columns wide and is terminated by an end-of-line indicator. The first six columns of every line contain a "record name". The first line holds the "HEADER" record which informs about the molecule type: either DNA or protein. Then, each polymer is described by a list of "ATOM" records; the list is terminated by a "TER" record. Every "ATOM" record holds atom symbol, chain identifier, residue number, atom coordinates and additional information. For DNA structure, chain and residue are respectively equivalent to strand and nucleotide. The activity diagram

of the *PDB reader* is described in Fig. 3. The function is in charge of parsing the PDB file in order to collect the following information: element, chain ID, residue number and atom coordinates.

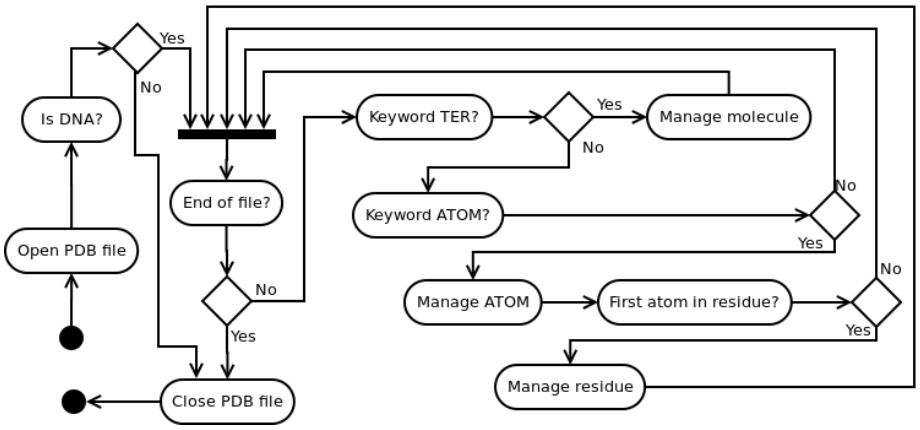


Fig. 3 : Activity diagram of the *PDB reader*. It reads any PDB file and converts it into Geant4 geometry.

This information is used to construct nested lists of geometry. The first list describes a molecule (for example a DNA strand). Each molecule is composed with a list of residues and each residue contains a list of atoms.

2.2. Geant4 geometry and visualization

In this part, we present the rendering of the PDB4DNA application applied to the "1ZBB PDB" file [29]. This file describes a tetranucleosome [30], a short complex of four nucleosomes. It has been selected for our application because it describes the largest DNA structure available in PDB format. The PDB4DNA application translates into Geant4 geometry the structure of the molecule only for visualization purposes by considering shapes, spatial transformations, materials and visualization attributes for each atom composing the molecule. We propose three different renderings using basic shapes like spheres and cylinders to model three different atomistic levels:

- The first rendering represents spheres centered on nucleotide barycenters (Fig. 4-a). The calculation of nucleotide barycenters (*PDBlib::computeNucleotideBarycenters*, Fig. 1) is also used to find the atoms nearest to each energy deposition (see Section 3.2).

- The second rendering proposes an atomistic view (Fig. 4-b). Spheres with Van der Waals radii are used to draw each atom. Corey-Pauling-Koltun (CPK) coloring convention is used to distinguish the different chemical elements.

-The third rendering provides a representation of nucleotides (group of atoms composed of sugar, phosphate and base) using spheres centered on barycenters of nucleotides. Each nucleotide is linked with a cylinder (Fig. 4-c).

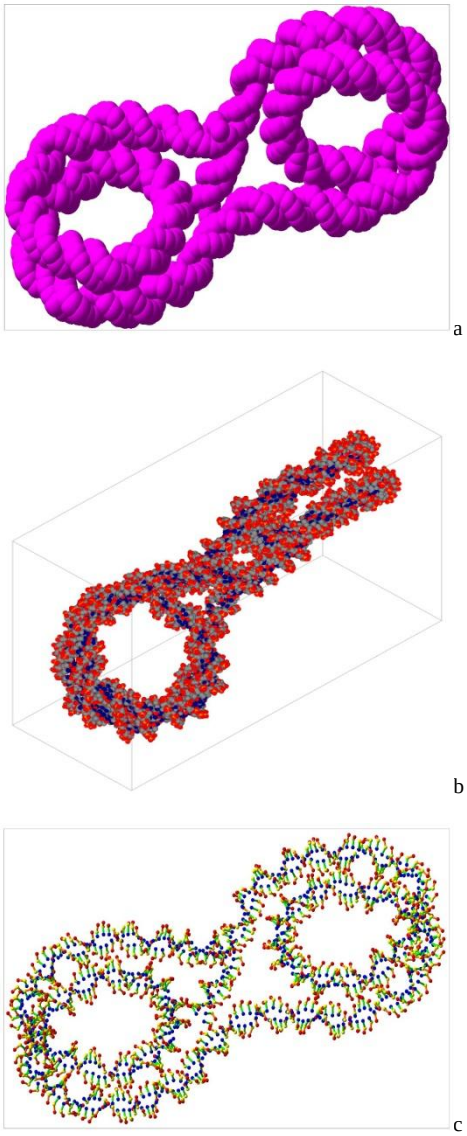


Fig. 4 : Rendering of a dinucleosome using the PDB4DNA application: a) spheres centered on nucleotide barycenters b) atomistic view and c) spheres centered on barycenters of nucleotides.

3. Simulation

3.1. Physics and geometry

PDB4DNA makes use of the Geant4-DNA physics list adapted to micro and nanodosimetry simulations [21]. Geant4-DNA processes extend the Geant4 electromagnetic physics for electrons, hydrogen and helium atoms (including charge states) and few ions, in liquid water down to very low energies. All interactions are explicitly simulated and energy deposition patterns are described down to

the nanometer scale. Since the Geant4-DNA toolkit does not yet propose cross sections for DNA compounds, simulations of particle tracks are performed in liquid water. The resulting energy depositions are then allocated to groups of atoms constituting the DNA molecule (section 3.2). To that purpose Geant4-DNA simulations are run inside a bounding box of liquid water having dimensions corresponding to the DNA molecule dimensions (*PDBlib::computeBoundingVolumeParams*, Fig. 1). Particles are generated from the box edges and randomly directed towards the box. Taking into account the effects of water radiolysis will be possible thanks to the integration of a specific module for radiation chemistry of Geant4-DNA integrated into the Geant4 10.1 release [22,23]. Our tracking approach is fully compatible since the production, diffusion and reaction of chemical species will be performed in a continuous medium made of liquid water.

3.2. Algorithm for finding the closest atom to energy depositions

Fig. 5 presents the activity diagram of the algorithm in charge of finding the closest atom to each energy deposition (*PDBlib::computeMatchEdepDNA*, Fig. 1). This algorithm has been optimized for DNA molecule and is not proposed for other stereochemistry conformation. First, the algorithm tries to allocate the energy deposition to a sphere bounding a nucleotide. The center of a bounding sphere is computed as the geometrical barycenter of a nucleotide. The radius is the maximum distance between the barycenter and atom coordinates constituting the nucleotide including the maximum Van der Waals radius (1.8 Angstrom for phosphor element). If an energy deposition is allocated to a bounding sphere, a second process checks Van der Waals radii to find the atom constituting the corresponding nucleotide nearest to the energy deposition. As nucleotide bounding spheres overlap, the two closest nucleotides are included in the algorithm. When a match is found, the algorithm returns the energy deposition, the DNA strand, the nucleotide identifier and the group identifier (base, phosphate or sugar group).

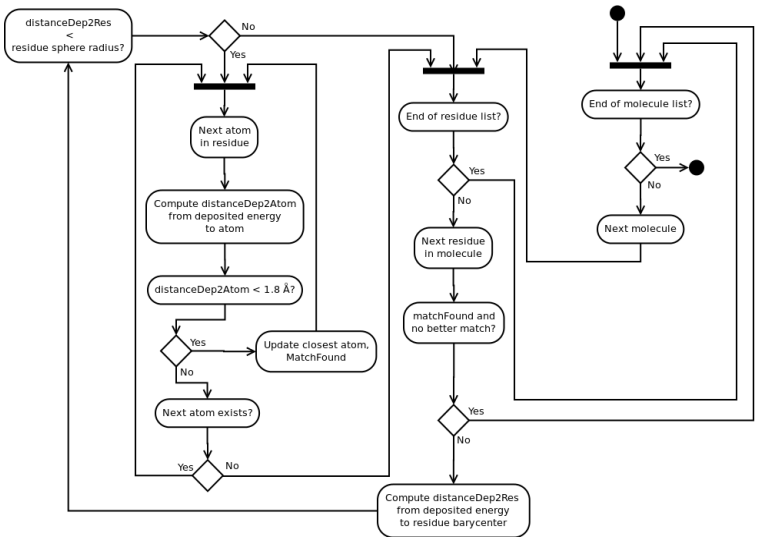


Fig. 5 : Activity diagram of the algorithm for searching the closest atom to energy deposition. This algorithm is optimized for DNA molecules and is not adapted for any other type of molecules.

3.3. Evaluation of DNA strand breaks

To compute DNA Single Strand Breaks (SSB) the assumption is made that a minimum energy deposition of 8.22 eV in a sugar-phosphate group induces a SSB. This energy threshold is chosen because it corresponds to the first excitation energy level of liquid water in Geant4-DNA models. This threshold value is adjustable in the simulation macro file. For DNA Double Strand Breaks (DSB) estimation, we assume that a maximum distance of 10 base pairs separating two SSBs on opposite DNA strands induces a DSB [7]; this default parameter is also adjustable. It should be noted that estimation of DSBs is supplied as demonstration only; the scientific relevance of DSBs in aqueous environment requires taking into account the indirect effects of ionizing radiation. At the beginning of the simulation, an empty associative map is created for each DNA strand. Key values of the map correspond to nucleotide identifiers and mapped values correspond to total energy deposition for each event which corresponds to the tracking of one primary particle and all its secondaries. Maps are updated each time the algorithm for finding the closest atom ends successfully. At the end of each event, maps are read to compute the number of SSBs and DSBs as presented in Fig. 6. When the simulation ends, total energy deposition in the bounding box, number of SSBs and number of DSBs are stored in a ROOT histogram [31]. As an example, Fig. 7-a illustrates the dinucleosome geometry extracted from the "1ZZB PDB" file while irradiated with monoenergetic protons whereas Fig. 7-b gives the estimation of SSBs and DSBs for different Linear Energy Transfer. The benchmarking of the

calculation of energy depositions into DNA geometry is being investigated in the framework of the Geant4-DNA project.

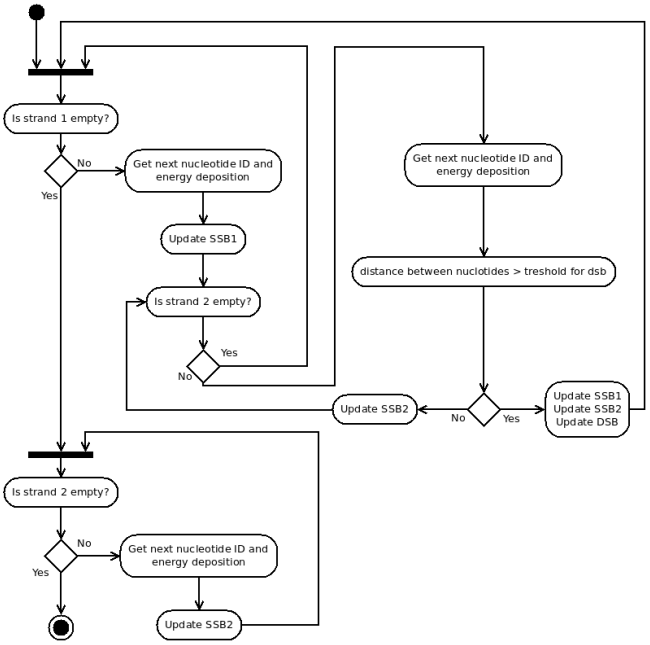
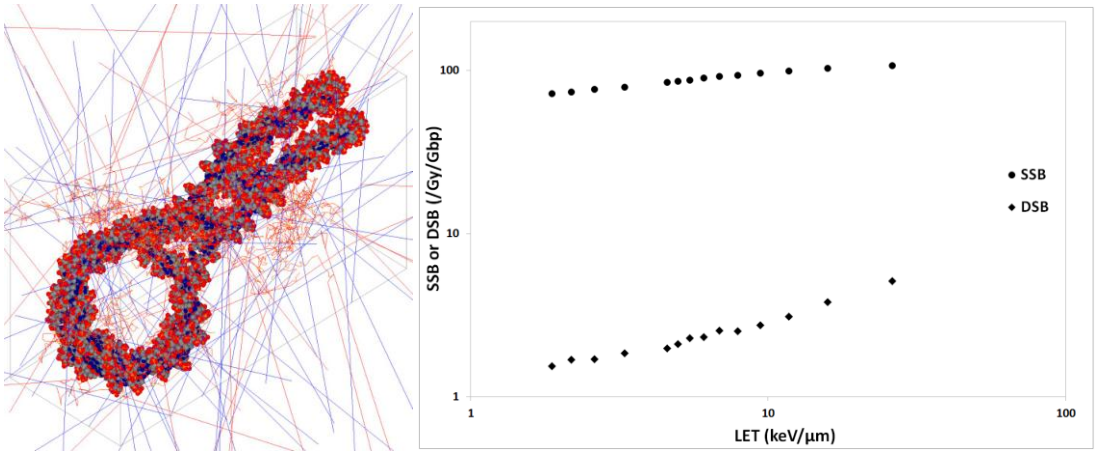


Fig. 6 : Activity diagram of the algorithm converting energy depositions into DNA strand breaks.



a) b)

Fig. 7 : a) Simulation of the irradiation of a dinucleosome (1ZBB PDB file) with monoenergetic protons and b) corresponding estimation of SSBs and DSBs as a function of Linear Energy Transfer (LET) of the monoenergetic protons.

4. Conclusion

The present paper highlights the functionalities of PDB4DNA, a new Geant4 user application dedicated to the estimation of direct DNA damage. PDB4DNA, based on the independent, cross-platform, free and open source C++ library PDBlib, is able to translate any PDB description of molecules into a Geant4 geometry down to the atomic level. In addition of visualization features, the

originality of PDB4DNA application comes from its ability to allocate very accurately each energy deposition to individual atom in the molecule during particle tracks. Then, results are automatically stored in a ROOT output file for further analysis by users. For the moment, calculations of energy depositions in DNA geometry are available for liquid water medium as cross sections for DNA material are not yet available in Geant4-DNA. In a near future, it is expected to enrich this application with the simulation of indirect effects through the developments proposed by the radiochemistry working-group of the Geant4-DNA collaboration. Effectively, one can assume that the algorithm proposed to search the nearest atom from particle energy deposition will be compliant with the evaluation of indirect effects produced by the diffusion of radiolysis products and their mutual interactions in a continuous water medium. Even though we demonstrate in this work our ability to easily model a simple DNA geometry (a dinucleosome) in order to calculate resultant energy depositions, we expect to propose in a future paper the modeling of a complete chromatin fiber using combinations of nucleosome geometries. In addition to SSB and DSB computation, the estimation of DNA fragment spectra is planned.

Acknowledgements

This work was partially supported by grants from Plan Cancer 2009-2013 French national initiative through the call for proposals “Domaine de la physique, des mathématiques ou des sciences de l’ingénieur appliqués au Cancer” managed by INSERM (Institut National de la Santé et de la Recherche Médicale) under contract PC201320.

Authors wish to thank the Geant4-DNA collaboration for its technical support, Michel Maire (Geant4 Associates International) for his advice concerning geometry transformations in Geant4, and Jean Orloff (Laboratoire de Physique Corpusculaire, Clermont-Ferrand) for his help concerning 3D rotation implementation.

References

[1] W. Friedland, M. Dingfelder, P. Kundrat, P. Jacob, Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC, *Mutat. Res.* **711** (2011) 28-40

[2] A.M. Kuzin, On the role of DNA in the radiation damage of the cell, *Int. J. Radiat. Biol.* **6** (1963) 201–209

[3] D.T. Goodhead, J. Thacker, R. Cox, Weiss lecture. Effects of radiations of different qualities on cells: molecular mechanisms of damage and repair, *Int. J. Radiat. Biol.* **63** (1993) 543–556

[4] P.F.R. Little, Structure and function of the human genome, *Genome Res.* **15** (2005) 1759-1766

[5] H. Nikjoo, S. Uehara, D. Emfietzoglou, F.A. Cucinotta, Track-structure codes in radiation research, *Radiat. Meas.* **41** (2006) 1052-1074

[6] D.E. Charlton, J.L. Humm, A method of calculating initial DNA strand breakage following the decay of incorporated 125I, *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* **53** (1988) 353–365

[7] H. Nikjoo, D.T. Goodhead, D.E. Charlton, H.G. Paretzke, Energy deposition in small cylindrical targets by ultrasoft X-rays, *Phys. Med. Biol.* **34** (1989) 691–705

[8] H. Nikjoo, D.T. Goodhead, D.E. Charlton, H.G. Paretzke, Energy deposition in small cylindrical targets by monoenergetic electrons, *Int. J. Radiat. Biol.* **60** (1991) 739–756

[9] M.A. Bernal and J.A. Liendo, An investigation on the capabilities of the PENELOPE MC code in nanodosimtery, *Med. Phys.* **36** (2009) 620-25

[10] S. Incerti, C. Champion, H.N. Tran, M. Karamitros, M.A. Bernal, Z. Francis, V. Ivanchenko, A. Mantero, Energy deposition in small-scale targets of liquid water using the very low energy electromagnetic physics processes of the Geant4 toolkit, *Nucl. Instrum. Methods Phys. Res. Sect. B* **306** (2013) 158-164.

[11] M. Dos Santos, C. Villagrasa, I. Clairand, S. Incerti, Influence of the DNA density on the number of clustered damages created by protons of different energies, *Nucl. Instrum. Methods Phys. Res. Sect. B* **298** (2013) 47-54

[12] V. Michalik and M. Begusova, Target model of nucleosome particle for track structure calculations and DNA damage modelling, *Int. J. Radiat. Biol.* **66** (1994) 267-277

[13] H. Nikjoo, P. Girard, A model of the cell nucleus for DNA damage calculations, *Int. J. Radiat. Biol.* **88** (2011) 87-97

[14] W. Friedland, P. Jacob, H.G. Paretzke, T. Stork, Monte Carlo simulation of the production of short DNA fragments by low linear energy transfer radiation using higher order DNA models, *Radiat. Res.* **150** (1998) 170-182

[15] V. Stepan, M. Davidkova, RADAMOL tool: Role of radiation quality and charge transfer in damage distribution along DNA oligomer, *Eur. Phys. J. D (2014)* doi:10.1140/epjd/e2014-50068-8

[16] B. Grosswendt and S. Pszona, The track structure of alpha particles from the point of view of ionization cluster formation in "nanometric" volumes of nitrogen particles, *Radiat. Environ. Biophys.* **41** (2002) 91-102

[17] R.E. Dickerson, H.R. Drew, B.M. Conner, R.M. Wing, A.V. Fratini, and M. L. Kopka, The anatomy of a-, b-, and z-DNA, *Science* (1982) **216** 475-485

[18] M.A. Bernal, C.E. de Almeida, C. Sampaio, S. Incerti, C. Champion, P. Nieminen, The invariance of the total direct DNA strand break yield, *Med. Phys.* **38** (2011) 4147-4153

[19] S. Agostinelli *et al.*, Geant4-a simulation toolkit, *Nucl. Instrum. Methods Phys. Res. Sect. A* **506** (2003) 250-303

[20] http://geant4-dna.org, accessed September 2014

[21] S. Incerti, A. Ivanchenko, M. Karamitros, A. Mantero, P. Moretto, H.N. Tran, B. Mascialino, C. Champion, V. Ivanchenko, M.A. Bernal, Z. Francis, C. Villagrasa, G. Baldacchino, P. Guèye, R. Capra, P. Nieminen, C. Zacharatou, Comparison of Geant4 very low energy cross section models with experimental data in liquid water, *Med. Phys.* **37** (2010) 4692-4708

[22] M. Karamitros, A. Mantero, S. Incerti, W. Friedland, G. Baldacchino, P. Barberet, M.A. Bernal, R. Capra, C. Champion, Z. El Bitar, Z. Francis, P. Guèye, A. Ivanchenko, V. Ivanchenko, H. Kurashige, B. Mascialino, P. Moretto, P. Nieminen, G. Santin, H. Seznec, H.N. Tran, C. Villagrasa, C. Zacharatou, Modeling Radiation Chemistry in the Geant4 Toolkit, *Prog. Nucl. Sci. and Technol.* **2** (2011) 503-508

[23] M. Karamitros, S. Luan, M.A. Bernal, J. Allison, G. Baldacchino, M. Davidkova, Z. Francis, W. Friedland, V. Ivantchenko, A. Ivantchenko, A. Mantero, P. Nieminem, G. Santin, H.N. Tran, V. Stepan, S. Incerti (2014) Diffusion-controlled reactions modeling in Geant4-DNA, *J. Comput. Phys.* **274** (2014) 841-882

[24] M.A. Bernal, D. Sikansi, F. Cavalcante, S. Incerti, C. Champion, V. Ivanchenko, Z. Francis, An atomistic geometrical model of the B-DNA configuration for DNA-radiation interaction simulations, *Comput. Phys. Commun.* **184** (2013) 2840-2847

[25] F.C. Bernstein, T.F. Koetzle, G.J. Williams, Jr. E.E. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The Protein Data Bank: A Computer-based Archival File For Macromolecular Structures, *J. Mol. Biol.* **112** (1977) 535-542

[26] www.salilab.org/~drussel/pdb/, accessed September 2014

[27] S. Loriot, F. Cazals, J. Bernauer, ESBTL: efficient PDB parser and data structure for the structural and geometric analysis of biological macromolecules, *Bioinformatics* **26** (2010) 1127-1128

[28] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, and G.R. Hutchison, Open Babel: An open chemical toolbox, *J. Cheminform.* **3** (2011) 33. See also http://openbabel.org, accessed September 2014

[29] doi: 10.2210/pdb1zbb/pdb

[30] T. Schalch, S. Duda, D.F. Sargent, T.J. Richmond, X-ray structure of a tetranucleosome and its implications for the chromatine fibre, *Nature* **436** (2005) 138-141

[31] R. Brun and F. Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, *Nucl. Instrum. Methods Phys. Res. Sect. A* **389** (1997) 81-86. See also http://root.cern.ch/, accessed September 2014

Figures

Fig. 1 : UML diagram of the PDB4DNA Geant4 user application: Geant4 virtual classes (white), Geant4 implemented classes (green), PDB library (purple) and interface to ROOT analysis software (red).

Fig. 2 : PDB file format description. This format provides atom coordinates for each nucleotide number (dark blue, light blue, purple, yellow) constituting each DNA strand (red and green).

Fig. 3 : Activity diagram of the *PDB reader*. It reads any PDB file and converts it into Geant4 geometry.

Fig. 4 : Rendering of a dinucleosome using the PDB4DNA application: a) spheres centered on nucleotide barycenters b) atomistic view and c) spheres centered on barycenters of nucleotides.

Fig. 5 : Activity diagram of the algorithm for searching the closest atom to energy deposition. This algorithm is optimized for DNA molecules and is not adapted for any other type of molecules.

Fig. 6 : Activity diagram of the algorithm converting energy depositions into DNA strand breaks.

Fig. 7 : a) Simulation of the irradiation of a dinucleosome (1ZBB PDB file) with monoenergetic protons and b) corresponding estimation of SSBs and DSBs as a function of Linear Energy Transfer (LET) of the monoenergetic protons.